Sharing Data from Large-Scale Genomic Studies

1) Legal/regulatory impediments to sharing data

As genome sequence becomes one of the principle standard data type for biological sciences concerns over privacy are becoming substantially increased. In the cancer space this is leading to segmentation of the research space. It is critical for the scientific community to exert itself on the political and legal frameworks to ensure that access to the data that are the life-blood of research continues. What steps might the US/UK engage in to develop effective solutions that support science and the development of therapies, while protecting patients?

2) Cloud sharing/computing

The scale of genome datasets continues to increase exponentially. For many researchers gaining access to data (when they can get through the legal/regulatory impediments) is becoming prohibitive because they cannot move the data effectively.

3) Quality of data and structure annotation

Recent discussions have highlighted the idea of “data brokers” groups that would serve to help data submitters accurately describe their research and communicate the data to archive repositories. Is there an effective pilot data broker that could be developed for cancer genome data?

4) Development of a cancer cell line compendium database

Major efforts in the US and UK are underway to characterize cancer cell lines. A unified vision for sharing these data to the public is necessary to prevent bifurcation of a critical resource.